

from the stored mails Mail 1 to Mail 4 are a building, a bill collecting, a customer, a bank and an account, and the like. If a word is extracted from the contents of the respective mails, "1" is given as the index value of the word. Otherwise, "0" is given as its index value. As a result, in table 1, it can be predicted that Tom is involved in bill collecting at the bank.

In this specification, a training example is presented by a set of attributes and values, and the result is given by a set of an attribute and a value. The cases shown in table 1 will be discussed as a training example. In the table 1, a building, a bill collecting, a customer, a bank and an account are the attributes of the problem, and the recipients are the attributes of the result. The learning agent 220 performs a machine learning for positive examples Mail 1 and Mail 2 of which recipient is Tom and negative examples Mail 3 and Mail 4 of which recipient is not Tom.

The learning result is described by using a decision tree. Each node of the decision tree represents a test. When a new problem is applied to this decision tree, the branches of the decision tree are traced according to the test result until the leaf node, where the solution is described, is reached.

The learning algorithm, e.g., ID3, is used to build the decision tree. The details of ID3 is described in "C4.5: Programs for Machine learning" by Quinlan, J.R.,

Morgan Kauffman, 1993. In the following, a simplified algorithm will be explained for the exemplary case shown in Table 1. Given a set of non-categorical attributes R, e.g., a building, a bill collecting, a customer, a bank and an account, a categorical attribute C, e.g., recipient, and a training data T, e.g., a set of mails, the decision tree is generated as follows:

function ID3 (R: a set of non-categorical attributes,
C: the categorical attribute,
T: a training set) returns a decision tree;

begin

If T is empty, return a single node with value
Failure;

If T consists of records with all of a same value for
the categorical attribute, return a single node with
that value;

If R is empty, then return, as a value, a single
node with the most frequent value among the values of
the categorical attribute that are found in records of
T;

Let A be the word with largest Gain(T,A) among
attributes in R;

Let $\{a_j \mid j=1,2,\dots,m\}$ be the values of attribute A;

Let $\{T_j \mid j=1,2,\dots,m\}$ be the subsets of T consisting

